DOCUMENT RESUME

ED 311 073 TM 013 910

AUTHOR Nevo, David
TITLE A Structured Method for Direct Assessment of Writing.
PUB DATE Mar 89
NOTE 15p.; Paper presented at the Annual Meeting of the American Educational Research Association (San Francisco, CA, March 27-31, 1989).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Elementary School Students; Essay Tests; Expressive Language; Foreign Countries; *Grade 6; Hebrew; Intermediate Grades; National Programs; National Surveys; Scores; Scoring Formulas; *Standardized Tests; *Test Construction; Test Validity; *Writing Evaluation
IDENTIFIERS *Direct Assessment; *Israel

ABSTRACT

The purpose of this study was to develop a testing method for the assessment of various types of writing at the elementary school level that would meet acceptable standards for educational measurement instruments as well as standards of utility and feasibility within a given educational system. The study was conducted within the framework of an Israeli national survey of sixth graders' writing performance. The direct assessment method chosen involved a variety of writing tasks, including school newsletter announcements, letters, and explanatory essays. The method distinguished among practical, expressive, and school writing. Five writing tasks were developed and field tested for each type of writing. The pilot study also included classroom observation and interviews with teachers and students. Scoring guides were developed for each writing task. The tasks were administered to a nationally representative sample of 2,590 sixth graders enrolled in a total of 57 schools. Approximately 800 responses were received for each writing task. A team of 12 lay scorers was trained to conduct the scoring. Data are now being analyzed to determine interrelationships between writing types and scoring methods and between test scores and other variables. Preliminary results indicate that the method may have curricular validity but not instructional validity. (TJH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED311073

A STRUCTURED METHOD FOR DIRECT ASSESSMENT OF WRITING

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

DAVID NEVO

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

DAVID NEVO

Tel Aviv University

Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, California, March 27-31, 1989

TM013910

BEST COPY AVAILABLE

The purpose of this study was to develop a testing method for the assessment of various types of writing at the elementary school level that would meet acceptable standards for educational measurement instruments as well as standards of utility and feasibility within a given educational system. The study was conducted within the framework of a national survey of student performance of writing in the sixth grade of elementary schools in Israel. The national survey was initiated by the Ministry of Education to study the instruction of writing in the elementary school, assess student writing performance at the end of elementary school, and develop recommendations for teachers and curriculum developers.

Considering the advantages and limitations of direct and indirect methods for the assessment of writing, a decision was made to prefer a direct assessment method with an attempt to assure reliability levels higher than previously reported for direct methods, and validity levels higher than those reported for indirect methods. Such a method, the Structured Writing Tasks (SWT) method, is presented in this paper with a rationale for its development, a description of the process of its development, and some data regarding its reliability, validity and utility.

The debate over using writing samples ("direct assessment") versus objective test items ("indirect assessment") to measure

writing ability is not new. The advantages and limitations of both methods have been widely discussed in the literature (e.g. Breland and Gaynor, 1979; Faigley et al., 1985; Stiggins, 1982). Defenders of indirect assessment methods point to their higher reliability as well as their concurrent and predictive validity. Indirect methods "score" also high on feasibility standards because of their low cost and simple administration and scoring. But school teachers seem to be skeptical of their validity and so are linguists and language testers who are concerned with their limited construct validity (Quellmalz et al., 1982). It is not clear what is actually being tested by indirect methods, nor is it clear what should be tested by writing tests.

On the other hand, defenders of direct assessment methods suggest that their reliability can be improved by structuring their procedures. They also suggest that direct assessment might have a better chance in obtaining content and construct validity if based on a sound conceptualization of writing and research findings. Various classifications of types of writing have been discussed and criticized in the literature (Applebee, 1981; Vahapassi, 1982; Quellmalz et al., 1982) but no agreement has been reached regarding the best way to classify types of writing. What seems to be agreed upon is the notion that the assessment of writing should deal with a wide range of writing types

representing the various kinds of writing that people use in their personal and professional life. In the SWT method we decided to use a classification which seemed to be communicative to teachers and curriculum developers and in accordance with the official curriculum determined by the Ministry of Education.

Concern for the purpose of writing and audience awareness are suggested by the literature as an important components in the planning of writing and its performance (Flower and Hayes, 1980; Odell and Goswami, 1982). The need to teach students how to write for real audiences and for specified purposes seems to be an important conclusion from the research findings of some studies concerned with what is called the rhetorical situation of writing. To us it suggested the need to specify in our writing tasks the purpose of writing and the audience that has to be addressed.

Writers' knowledge of subject matter may have a considerable influence on how well they write on a certain subject. Research on writing suggests that topic knowledge has an influence on writing performance (Quellmalz et al., 1980). Thus, in structuring writing tasks for the assessment of writing ability there is a need to control the influence of topic knowledge by means of topic selection or by provision of information within the test.

Developments in the field of educational evaluation (Nevo, 1983) such as the distinction between criterion-referenced and

norm-referenced test (Glaser, 1963), formative and summative functions of assessment (Scriven, 1967) and the distinction between description and judgment (Stake, 1967) should also be considered when developing an assessment system or using its products.

Overall, the SWT method can be characterized as adhering to the following principles:

(a) In the assessment of writing a distinction should be made between various types of writing. In our case we made a distinction between practical writing, expressive writing and school writing.

(b) For each type of writing authentic writing tasks have to be developed according to the educational and social context of the target population to be assessed. We made an attempt to draw our writing tasks from the world of sixth grade students.

(c) In developing writing tasks topic knowledge should be controlled by providing necessary information to the writer or assuming its existence. Some writing tasks were selected on the assumption that all six graders possess the knowledge necessary to respond to such tasks. For other tasks the necessary information was provided within the framework of the test.

(d) Each writing task should identify the audience and the purpose of the writing. We did so for the writing tasks related to practical and expressive writing but not to school writing. School writing in our educational system is not very much audience oriented and students rarely write for anyone except their teachers.

(e) Multiple scoring procedures should be developed for each writing task according to the function of the assessment and its potential use. We developed four scoring methods for each writing task: holistic-norm referenced, holistic-criterion referenced, analytic-norm referenced, and analytic-criterion referenced.

(f) In each scoring procedure four components should be considered: content, structure, language and mechanics. The weight of each component in a composite score (if such a score is necessary), should be determined according to the function of the assessment and on the basis of research findings.

(g) Specific scoring guides should be developed for each writing task relating to the particular purpose and specified audience of each task. Following the advice of Primary-Trait Scoring, developed by Lloyd-Jones (1977) and used in the National Assessment of Educational

Progress, we believe that such specific scoring guides if used by teachers could weaken their tendency to score writing samples mainly on general criteria such as grammar, vocabulary or spelling.

DEVELOPMENT OF THE SWT METHOD

The SWT method was developed in a systematic step by step process and in close cooperation between the research team and an active steering committee comprised of teachers, school supervisors, curriculum developers and linguists from the academe. The function of the steering committee was threefold: to make policy decisions regarding the development of assessment method and the conduct of the national writing survey, to secure school cooperation with the study, and to facilitate potential utilization of study results.

At the **first stage** we analyzed the national curriculum for language instruction in the elementary school and inspected a sample of Hebrew textbooks and other instructional materials. We also reviewed at this stage current literature on writing research and literature on the assessment of writing. At the end of this stage a decision was made to distinguish between three types of writing: practical writing, expressive writing and school writing.

At the second stage several writing tasks were developed for each type of writing according to the principles mentioned in the previous section. The tasks were then presented to the steering committee and revised on the basis of their comments. The following are examples of writing tasks for the various types of writing:

Practical writing: As a member of your class board you suggested to your school principal to open a computer club in the school. The principal accepted the idea on the condition that at least 30 students will participate in the club.
Write an announcement for the school newsletter in which you give the necessary details regarding the computer club, try to convince students to join, and explain registration procedures.

Expressive writing: Your best friend has left to live in another town and you are very sad about it.
Write to him a letter in which you describe how you feel and how much you miss him.

School Writing: **Explain why we celebrate the holiday of Hannuka, and describe the customs related to this holiday.**

At the third stage five writing tasks for each type of writing were field tested in a pilot study conducted in 15 classes. In addition to test administration the pilot study also included classroom observations and interviews with teachers and students.

At the fourth stage scoring guides were developed for each writing task. The scoring procedures were tried out on a sample of

writings from the pilot study. As a result of this tryout a decision was made to choose two writing tasks for each type of writing and to use four scoring methods for each task. The scoring methods were: a holistic-norm referenced method, a holistic-criterion referenced method, an analytic-norm referenced method, and an analytic-criterion referenced method.

At the **fifth stage** the writing tasks were administered to a nationally representative sample of 2590 sixth grade students studying in 96 classes within 57 schools. Each student wrote on two writing tasks, that were assigned to him on a random basis, and answered a short questionnaire regarding the test. Thus, for each writing task about 800 responses were obtained. Data were also collected on students' school grades, and teacher questionnaires were administered in participating classes regarding writing instruction and testing practice.

At the **sixth stage** a team of 12 lay scorers were trained to conduct the scoring. A special procedure of monitoring individual scorers' reliability was used during intensive scoring sessions, and corrections of scoring were introduced whenever it was necessary, to assure an overall high level of reliability.

At the final stage of the study, now still being completed, data were analyzed to provide information regarding interrelations among various types of writing and among various scoring methods, as well as relationships between test scores and other variables.

SOME FINDINGS

Although the data of the study are still being analyzed and the findings of the national survey have not been published yet, some preliminary findings regarding the SWT as an assessment method can be mentioned at the present time.

Qualitative data regarding content validity resulting from the curriculum, conducted during the first stage of the study, and descriptive data obtained from teacher questionnaires regarding writing instruction, suggest that the SWT as developed in this study might have curricular validity but not instructional validity. Within the framework of the Israeli elementary school the three types of writing assessed in our study (practical, expressive and school writing) seem to represent the official curriculum but not necessarily what is being taught in school, if writing is being taught at all.

The reliability findings already obtained for the various scoring procedures of the SWT method are quite encouraging. For the holistic criterion-referenced scoring procedure the following interrater reliability coefficients have been obtained: $r = .92$ for practical writing, $r = .89$ for expressive writing, and $r = .84$ for school writing. Similar findings were obtained for other scoring procedures. These findings suggest that a direct assessment method of writing performance can reach high levels of interrater reliability if it is structured in a systematic way and carefully implemented.

The distinction between holistic scoring and analytic scoring seems to be an important one. We have found correlations of about r = .50 between holistic-norm referenced and holistic-criterion referenced scores, but very low correlations between holistic scores and composite scores obtained by simple means of sub-scores for the various components comprising the analytic scoring.

Preliminary analyses of holistic scores and sub-scores in the analytic scoring procedures revealed considerable differences among the various types of writing indicating differences in the relative importance of components such as content, structure, language and mechanics, in the assessment of various types of writing. As an example, mechanics seems to be weighted high in practical writing but not in expressive writing and vice versa for content. More analysis is needed in this regard before further recommendations can be made regarding the weighting of such analytic components when writing assessment is used to serve various educational functions.

REFERENCES

Applebee, A.N. (1981). Writing in the secondary school: English and the content areas. Urbana, IL:NCTE.

Breland, H.M. & Gaynor, J.L. (1979). A comparison of direct and indirect assessments of writing skill. Journal of Educational Measurement, 16, 119-128.

Faigley, L., Cherry, R.D., Jolliffe, D.A. & Skinner, A.M. (1985). Assessing writers' knowledge and process of composing. Norwood, NJ: Ablex.

Flower, L. & Hayes, J.R. (1980). The cognition of discovery: Defining a rhetorical problem. College Composition and Communication, 31, 21-32.

Glaser, R. (1963). Instructional technology and the measurement of learning outcomes: Some questions. American Psychologist, 18, 519-521.

Lloyd-Jones, R. (1977). Primary trait scoring. In C.R. Cooper, & L. Odell (Eds.), Evaluating writing: Describing, measuring, judging. Urbana, IL:NCTE.

Nevo, D. (1983). The conceptualization of educational evaluation: An analytical review of the literature. Review of Educational Research, 53, 117-128.

Odell, L., & Goswani, D. (1982). Writing in a non-academic setting. Research in the Teaching of English, 16, 201-223.

Quellmalz, E.S., Baker, E.L., & Enright, G. (1980). Test design: A comparison of modalities of writing prompts. Los Angeles: UCLA Graduate School of Education.

Quellmalz, E.S., Capell, F., & Chou, C. (1982). Effects of discourse and response mode on the measurement of writing competence. Journal of Educational Measurement, 19, 241-258.

Scriven, M. (1967). The methodology of evaluation. In R.E. Stake (Ed.), AERA monograph series on curriculum evaluation, No. 1. Chicago: Rand McNally.

Stake, R.E. (1967). The countenance of educational evaluation. Teachers College Record, 68, 523-540.

Stiggins, R.J. (1982). A comparison of direct and indirect writing assessment methods. Research in the Teaching of English, 16, 101-114.

Vahapassi, A. (1982). On the specification of the domain of school writing. Evaluation in Education, 5, 265-290.